Cal_data_format deming

Load Libraries into R Environment

```
library(lubridate)
```

```
Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v forcats 1.0.0     v stringr 1.5.1
v ggplot2 3.5.1     v tibble  3.2.1
v purrr   1.0.2     v tidyr   1.3.1
v readr   2.1.5

-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(RODBC)
library(reshape2)
```

```
Attaching package: 'reshape2'

The following object is masked from 'package:tidyr':

    smiths
```

```
library(ggplot2)
library(GGally)
```

```
Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2
```

Read in sensor data

```
#Read in each day's data as separate dataframe per sensor.
wb <- "C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/PA Test Combined Deming-2024
con2 <- odbcConnectExcel2007(wb)
day1 <- sqlFetch(con2, "20240311")
day2 <- sqlFetch(con2, "20240312")
day3 <- sqlFetch(con2, "20240313")
day4 <- sqlFetch(con2, "20240314")
day5 <- sqlFetch(con2, "20240315")
day6 <- sqlFetch(con2, "20240316")
```

```
day7 <- sqlFetch(con2, "20240317")
day8 <- sqlFetch(con2, "20240318")
```

For each sensor, merge daily dataframes into single dataframe

```
gc9 <- rbind(day1,day2,day3,day4,day5,day6,day7,day8)
```

Write combined sensor dataframe to csv

```
write.csv(gc9, "C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/GC
```

Subset for data we want (time and pm2.5 concentration)

```
gc9f <- gc9[c("UTCDateTime", "pm2_5_atm")]
```

Format column names

```
colnames(gc9f) <- c("time", "GC9_PM2_5")
```

Write formatted sensor dataframe to csv

```
write.csv(gc9f, "C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/G
```

Generate time sequence

```
#create time sequence of every second from 2024-03-11 to 2024-03-19
time_sequence <- seq(
  from = as.POSIXct("2024-03-11 00:00:00", tz = "UTC"),
  to = as.POSIXct("2024-03-19 00:00:00", tz = "UTC"),
  by = "sec"
)
```

Turn time sequence into dataframe

```
time_sequence_df <- as.data.frame(time_sequence)
```

Read in formatted sensor data

```
GC1 <- read.csv("C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/G
GC2 <- read.csv("C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/G
GC3 <- read.csv("C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/G
GC4 <- read.csv("C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/G
GC5 <- read.csv("C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/G
GC6 <- read.csv("C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/G
GC7 <- read.csv("C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/G
GC8 <- read.csv("C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/G
GC9 <- read.csv("C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/G
GC10 <- read.csv("C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/Deming_analysis/
```

Put sensor dataframes into list

```
my_data <- list(GC1, GC2, GC3, GC4, GC5, GC6, GC7, GC8, GC9, GC10)
```

Format column names so timestamp mataches

```
colnames(time_sequence_df) <- c("timestamp_datetime")
```

Format time stamps

```
#loop through list of sensor dataframes, format timestamp column from character to date-time
for (x in seq_along(my_data)) {
  df <- my_data[[x]]
  df$timestamp_clean <- sub("z$", "+0000", df$time)
  df$timestamp_datetime <- as.POSIXct(df$timestamp_clean, format = "%Y/%m/%dT%H:%M:%S%z", tz
  my_data[[x]] <- df

}
```

Join each sensor data to time series so all records are temporally concurrent

```
datacomb <- left_join(time_sequence_df, my_data[[1]], by="timestamp_datetime")
datacomb1 <- left_join(datacomb, my_data[[2]], by="timestamp_datetime")
datacomb2 <- left_join(datacomb1, my_data[[3]], by="timestamp_datetime")
datacomb3 <- left_join(datacomb2, my_data[[4]], by="timestamp_datetime")
datacomb4 <- left_join(datacomb3, my_data[[5]], by="timestamp_datetime")
datacomb5 <- left_join(datacomb4, my_data[[6]], by="timestamp_datetime")
datacomb6 <- left_join(datacomb5, my_data[[7]], by="timestamp_datetime")
datacomb7 <- left_join(datacomb6, my_data[[8]], by="timestamp_datetime")
datacomb8 <- left_join(datacomb7, my_data[[9]], by="timestamp_datetime")
datacomb9 <- left_join(datacomb8, my_data[[10]], by="timestamp_datetime")
```

Subset data for records we want (time, PM2.5 for each sensor)

```
datacomb_format <- subset(datacomb9, select = c(timestamp_datetime, GC1_PM2_5, GC2_PM2_5, GC3
```

Remove rows where there is no record from any sensor

```
datacomb_format_clean <- datacomb_format[!with(datacomb_format,is.na(GC1_PM2_5)& is.na(GC2_PM
```

Write cleaned and formatted data to csv

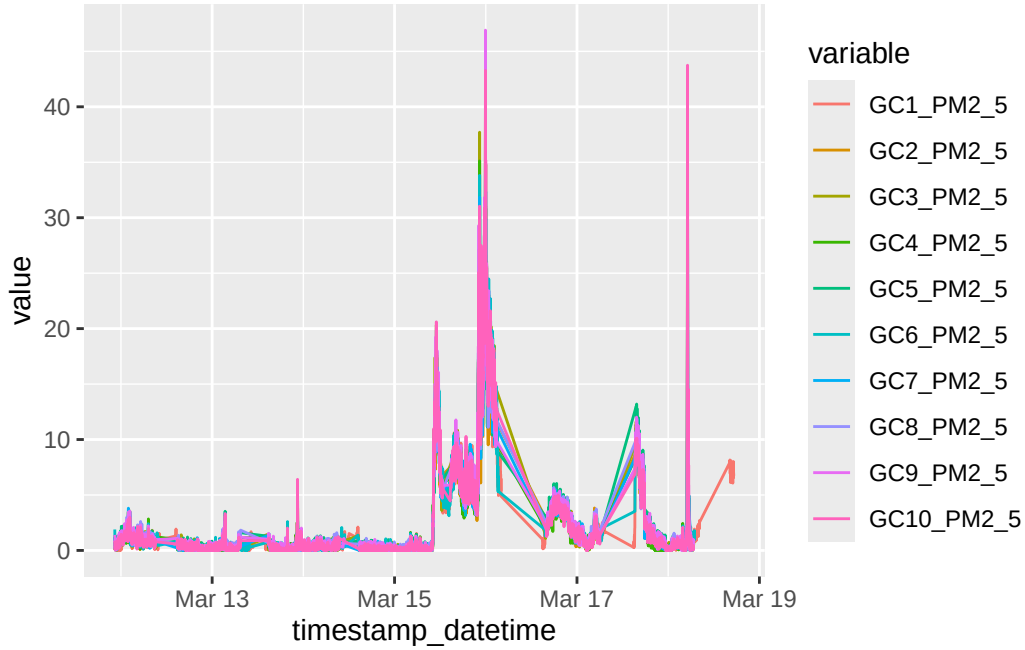
```
write.csv(datacomb_format_clean, "C:/Users/jacks/Documents/UNM/P30/AIRWISE/Calibration_data/M
```

Transpose data for plotting

```
#Transpose data to long format so we have 3 columnsfor plotting (time, sensor ID, and PM2.5)
datacomb_format_clean_long <- melt(datacomb_format_clean, id.vars = "timestamp_datetime", var
```

Plot data

```
ggplot(datacomb_format_clean_long, aes(timestamp_datetime, value, group = variable, color = v
  geom_line(data=datacomb_format_clean_long[!is.na(datacomb_format_clean_long$value),])
```



Plot indicates that there is slight variance, but overall strong agreement between sensors across time.

```
#Data summary
summary(datacomb_format_clean)
```

```
 timestamp_datetime                 GC1_PM2_5        GC2_PM2_5
 Min.   :2024-03-11 22:22:40.00   Min.   : 0.000   Min.   : 0.000
 1st Qu.:2024-03-13 15:41:37.00   1st Qu.: 0.120   1st Qu.: 0.100
 Median :2024-03-14 23:50:06.00   Median : 0.550   Median : 0.420
 Mean   :2024-03-15 01:07:49.86   Mean   : 2.194   Mean   : 1.834
 3rd Qu.:2024-03-16 18:21:20.25   3rd Qu.: 1.930   3rd Qu.: 1.450
 Max.   :2024-03-18 17:15:41.00   Max.   :40.220   Max.   :36.000
                                  NA's   :28156    NA's   :28910
   GC3_PM2_5        GC4_PM2_5        GC5_PM2_5        GC6_PM2_5
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.00
 1st Qu.: 0.140   1st Qu.: 0.128   1st Qu.: 0.260   1st Qu.: 0.21
 Median : 0.490   Median : 0.445   Median : 0.710   Median : 0.64
 Mean   : 2.296   Mean   : 2.119   Mean   : 2.266   Mean   : 2.37
```

```
 3rd Qu.: 1.712   3rd Qu.: 1.520   3rd Qu.: 2.020   3rd Qu.: 1.86
 Max.   :43.230   Max.   :38.880   Max.   :38.720   Max.   :38.95
 NA's   :28650    NA's   :28790    NA's   :28740    NA's   :28420
   GC7_PM2_5        GC8_PM2_5        GC9_PM2_5        GC10_PM2_5
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 0.120   1st Qu.: 0.240   1st Qu.: 0.220   1st Qu.: 0.150
 Median : 0.470   Median : 0.660   Median : 0.660   Median : 0.510
 Mean   : 1.975   Mean   : 2.163   Mean   : 2.460   Mean   : 2.265
 3rd Qu.: 1.530   3rd Qu.: 1.830   3rd Qu.: 2.087   3rd Qu.: 1.750
 Max.   :36.790   Max.   :39.910   Max.   :46.910   Max.   :43.750
 NA's   :28870    NA's   :28800    NA's   :28680    NA's   :28680
```

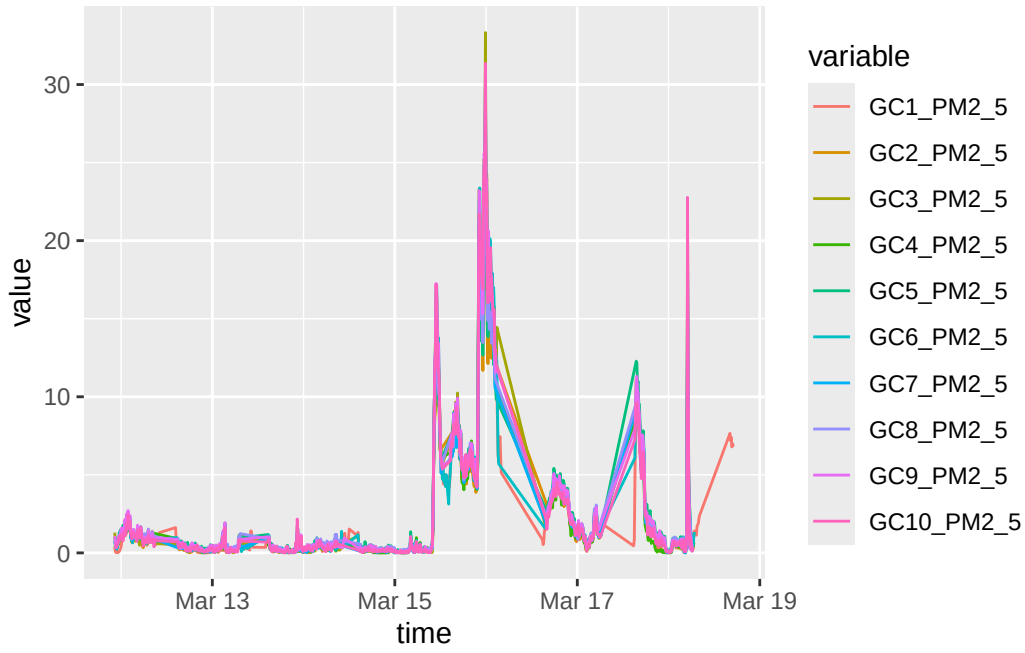
Slight variance in reported PM2.5 values. Relative strong agreeement in mean recorded PM2.5, with larger variance observed in max recorded value (+- 10 ug/m3 PM2.5).

Aggregate data to 10 minutes

```
datacomb_agg <- datacomb_format_clean %>%
  group_by(time = floor_date(timestamp_datetime, '10 minutes')) %>%
  summarize(GC1_PM2_5 = mean(GC1_PM2_5, na.rm = TRUE), GC2_PM2_5 = mean(GC2_PM2_5, na.rm = TR
            GC6_PM2_5 = mean(GC6_PM2_5, na.rm = TRUE), GC7_PM2_5 = mean(GC7_PM2_5, na.rm = TR
```

```
datacomb_agg_long <- melt(datacomb_agg, id.vars = "time", variable.name = "variable", value.n
```

```
ggplot(datacomb_agg_long, aes(time, value, group = variable, color = variable)) +
  geom_line(data=datacomb_agg_long[!is.na(datacomb_agg_long$value),])
```




```
#Print head of data table
head(datacomb_agg)
```

```
# A tibble: 6 x 11
  time                GC1_PM2_5 GC2_PM2_5 GC3_PM2_5 GC4_PM2_5 GC5_PM2_5
  <dttm>                  <dbl>     <dbl>     <dbl>     <dbl>     <dbl>
1 2024-03-11 22:20:00     0.328     0.285     1.30      0.478     0.985
2 2024-03-11 22:30:00     0.08      0.438     0.53      0.384     0.872
3 2024-03-11 22:40:00     0.018     0.426     0.504     0.376     0.612
4 2024-03-11 22:50:00     0.086     0.206     0.452     0.3       0.532
5 2024-03-11 23:00:00     0.01      0.402     0.58      0.552     0.868
6 2024-03-11 23:10:00     0.014     0.448     0.588     0.616     0.982
# i 5 more variables: GC6_PM2_5 <dbl>, GC7_PM2_5 <dbl>, GC8_PM2_5 <dbl>,
#   GC9_PM2_5 <dbl>, GC10_PM2_5 <dbl>
```

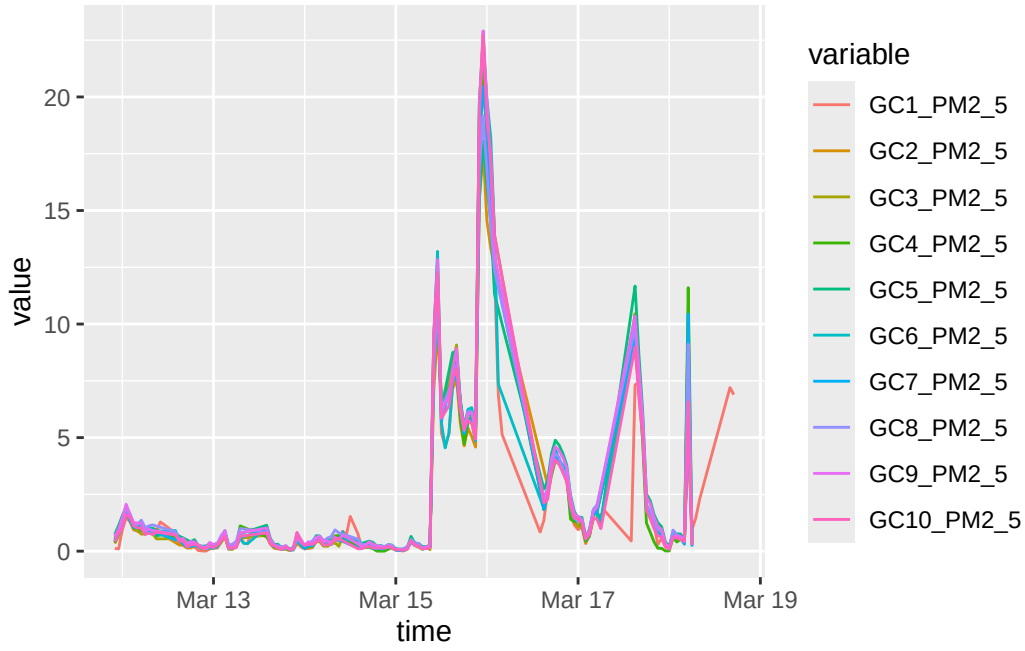
Note slight variance (+-1 ug/m3 PM2.5) between sensors.

Aggregate data to 1 hour average

```r
datacomb_agg_1hr <- datacomb_format_clean %>%
  group_by(time = floor_date(timestamp_datetime, '1 hour')) %>%
  summarize(GC1_PM2_5 = mean(GC1_PM2_5, na.rm = TRUE), GC2_PM2_5 = mean(GC2_PM2_5, na.rm = TR
            GC6_PM2_5 = mean(GC6_PM2_5, na.rm = TRUE), GC7_PM2_5 = mean(GC7_PM2_5, na.rm = TR
```

```r
datacomb_agg_long_1hr <- melt(datacomb_agg_1hr, id.vars = "time", variable.name = "variable"
```

```r
ggplot(datacomb_agg_long_1hr, aes(time, value, group = variable, color = variable)) +
  geom_line(data=datacomb_agg_long_1hr[!is.na(datacomb_agg_long_1hr$value),])
```



Test for sensor agreement through pairwise correlations

```r
vars <- c(
  "GC1_PM2_5", "GC2_PM2_5", "GC3_PM2_5", "GC4_PM2_5", "GC5_PM2_5", "GC6_PM2_5", "GC7_PM2_5",
)

datacomb_sub <- datacomb_agg[, c(vars)]

p_cor <- ggpairs(
```

```
  datacomb_sub,
  upper = list(continuous = wrap("points", alpha = 0.2, size = 0.5)),
  lower = list(continuous = "cor")
)

print(p_cor)
```

```
Warning: Removed 146 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 96 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 120 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 113 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 52 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 133 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 124 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 101 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 102 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 146 rows containing missing values

Warning: Removed 146 rows containing non-finite outside the scale range
(`stat_density()`).
```

```
Warning: Removed 146 rows containing missing values or values outside the scale range
(`geom_point()`).
Removed 146 rows containing missing values or values outside the scale range
(`geom_point()`).
Removed 146 rows containing missing values or values outside the scale range
(`geom_point()`).
Removed 146 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 153 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 146 rows containing missing values or values outside the scale range
(`geom_point()`).
Removed 146 rows containing missing values or values outside the scale range
(`geom_point()`).
Removed 146 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 96 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 146 rows containing missing values

Warning: Removed 96 rows containing non-finite outside the scale range
(`stat_density()`).

Warning: Removed 121 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 114 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 96 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 134 rows containing missing values or values outside the scale range
(`geom_point()`).
```

```
Warning: Removed 125 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 107 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 106 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 120 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 146 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 121 rows containing missing values

Warning: Removed 120 rows containing non-finite outside the scale range
(`stat_density()`).

Warning: Removed 122 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 120 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 133 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 125 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 121 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 120 rows containing missing values or values outside the scale range
(`geom_point()`).
```

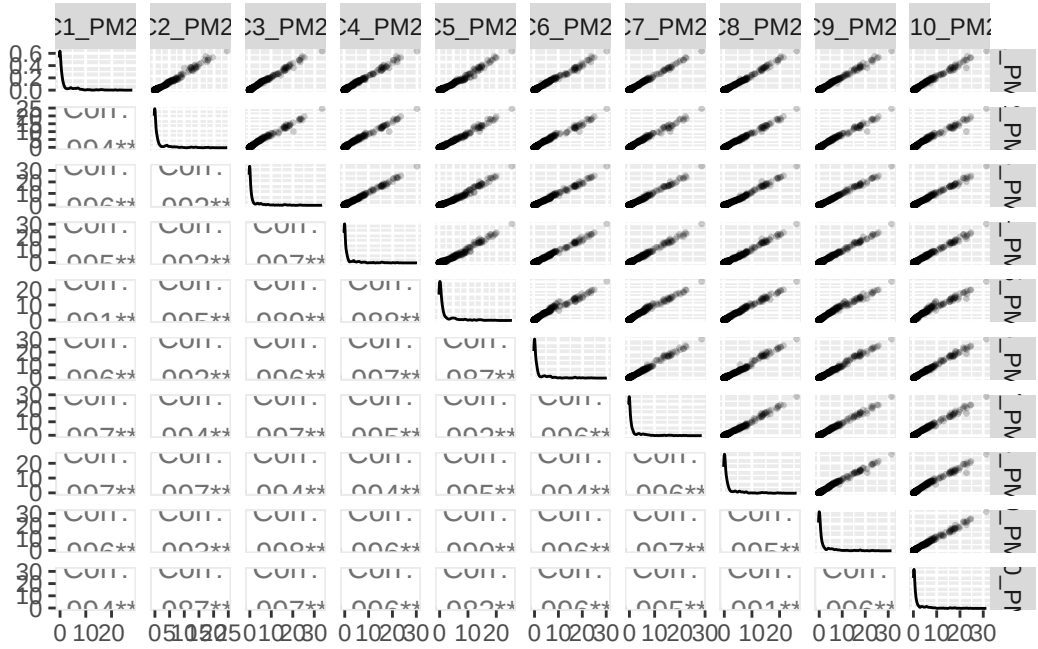
```
Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 113 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 146 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 114 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 122 rows containing missing values

Warning: Removed 113 rows containing non-finite outside the scale range
(`stat_density()`).

Warning: Removed 113 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 134 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 124 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 115 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 113 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 52 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 146 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 96 rows containing missing values
```

```
Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 120 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 113 rows containing missing values

Warning: Removed 52 rows containing non-finite outside the scale range
(`stat_density()`).

Warning: Removed 133 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 124 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 101 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 102 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 133 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 153 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 134 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 133 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 134 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 133 rows containing missing values
```

```
Warning: Removed 133 rows containing non-finite outside the scale range
(`stat_density()`).

Warning: Removed 134 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning: Removed 133 rows containing missing values or values outside the scale range
(`geom_point()`).
Removed 133 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 124 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 146 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 125 rows containing missing values
Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 125 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 124 rows containing missing values
Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 124 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 134 rows containing missing values

Warning: Removed 124 rows containing non-finite outside the scale range
(`stat_density()`).

Warning: Removed 124 rows containing missing values or values outside the scale range
(`geom_point()`).
Removed 124 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 101 rows containing missing values
```

```
Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 146 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 107 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 121 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 115 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 101 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 133 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 124 rows containing missing values

Warning: Removed 101 rows containing non-finite outside the scale range
(`stat_density()`).

Warning: Removed 105 rows containing missing values or values outside the scale range
(`geom_point()`).

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 102 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 146 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 106 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 120 rows containing missing values
```

```
Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 113 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 102 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 133 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 124 rows containing missing values

Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
Removed 105 rows containing missing values

Warning: Removed 102 rows containing non-finite outside the scale range
(`stat_density()`).
```



Pairwise correlation tests indicate high precision across sensors.